

# Competition and adaptation in an Internet evolution model

M<sup>a</sup> Ángeles Serrano, Marián Boguñá and Albert Díaz-Guilera<sup>1</sup>

<sup>1</sup>*Departament de Física Fonamental, Universitat de Barcelona, Av. Diagonal 647, 08028 Barcelona, Spain*

(Dated: February 2, 2008)

We model the evolution of the Internet at the Autonomous System level as a process of competition for users and adaptation of bandwidth capability. We find the exponent of the degree distribution as a simple function of the growth rates of the number of autonomous systems and the total number of connections in the Internet, both empirically measurable quantities. This fact place our model apart from others in which this exponent depends on parameters that need to be adjusted in a model dependent way. Our approach also accounts for a high level of clustering as well as degree-degree correlations, both with the same hierarchical structure present in the real Internet. Further, it also highlights the interplay between bandwidth, connectivity and traffic of the network.

PACS numbers: 89.75.-k, 87.23.Ge, 05.70.Ln

A statistical physics approach to Internet modeling will be successful only if its large-scale properties can be explained and predicted on the basis of the interactions between basic units at the microscopic level [1, 2]. Dynamical evolution rules acting at the local scale would then determine the behavior and the emergent structural properties of the whole Internet, which self-organizes under an absolute lack of centralized control [3, 4]. This approach is at the core of a set of recent network models focusing on evolution, which recognise growth as one of the key mechanisms on network formation, along with preferential attachment or other utility rules [5, 6, 7, 8, 9, 10]. While several of such models succeed in depicting some of the Internet features, none of them accounts for a complete description of the real topology [11, 12, 13]. In this paper, we present a new growing network model which, from competition and adaptation mechanisms, reproduces the topological properties observed in the autonomous system level maps of the Internet, namely: i) a scale-free distribution of the number of connections –or degree– of vertices  $k_i$ , characterized by a power law  $P(k) \sim k^{-\gamma}$ ,  $2.1 \leq \gamma < 2.5$ , ii) high clustering coefficient  $c_k$ , defined as the ratio between the number of connected neighbors of a node of degree  $k$  and the maximum possible value averaged for all nodes of degree  $k$ , and, finally, iii) disassortative degree-degree correlations, quantified by means of the average nearest neighbors degree of nodes of degree  $k$ ,  $\bar{k}_{nn}(k)$  [12].

We start our analysis by looking at the growth of the Internet during the last three decades. We focus on the temporal evolution of the number of hosts present in the Internet [14] as compared to the number of distinct autonomous systems (ASs) and the total number of connections among them. We have reanalysed AS maps collected by the *Oregon route-views* project which has recorded the Internet topology at the AS level since November 1997 [15]. Let  $W(t)$ ,  $N(t)$  and  $E(t)$  be the total number of hosts (we assume that number of hosts is equivalent to number of users), number of ASs and edges among ASs at time  $t$  respectively. Fig.1 shows

empirical measurements for these quantities revealing exponential growths,  $W(t) \sim W_0 e^{\alpha t}$ ,  $N(t) \sim N_0 e^{\beta t}$  and  $E(t) \sim E_0 e^{\delta t}$  respectively, with rates  $\alpha = 0.036$ ,  $\beta = 0.029$  and  $\delta = 0.033$ , where  $\alpha \gtrsim \delta \gtrsim \beta$ . These exponential growths, in turn, determine the scaling relations with the system size, that is,  $W \sim N^{\alpha/\beta}$ ,  $E \sim N^{\delta/\beta}$  and  $\langle k \rangle \sim N^{\delta/\beta-1}$  [16]. All three rates are, indeed, quite close to each other. This result poses the question of whether these inequalities actually hold or, in contrast, are due to statistical fluctuations. A simple argument will convince us that the inequalities are, actually, the natural answer. There are two mechanisms capable to compensate an increase in the number of users: the creation of new ASs and the creation of new connections by old ASs. When both mechanisms take place simultaneously, the rate of growth of new ASs,  $\beta$ , must necessarily be smaller than  $\alpha$ , whereas the rate of growth of the number of connections,  $\delta$ , must be greater than  $\beta$ . Any other situation would lead to an imbalance between the number of users and the maximum number of users that the system can manage.

Our model is defined according to the following rules: (i) At rate  $\alpha W(t)$ , new users join the system and choose provider  $i$  according to some preference function,  $\Pi_i(\{\omega_j(t)\})$ , where  $\omega_j(t)$ ,  $j = 1, \dots, N(t)$ , is the number of hosts connected to AS  $j$  at time  $t$ . The function  $\Pi_i(\{\omega_j(t)\})$  is normalised so that  $\sum_i \Pi_i(\{\omega_j(t)\}) = 1$  at any time. (ii) At rate  $\beta N(t)$ , new ASs join the network with an initial number of users,  $\omega_0$ , randomly withdrawn from the pool of users already attached to existing ASs. Therefore,  $\omega_0$  can be understood as the minimum number of users required to keep ASs in business. (iii) At rate  $\lambda$ , each user changes his provider and chooses a new one using the same preference function  $\Pi_i(\{\omega_j(t)\})$ . Finally, (iv) each node tries to adapt its number of connections to other nodes according to its present number of users, in an attempt to provide them an adequate access to the Internet. We will discuss this last point in the second part of the work. With the above ingredients, in the continuum approximation, the dynamics of single nodes is

described by the stochastic differential equation

$$\frac{d\omega_i}{dt} = A(\omega_i, t) + [D(\omega_i, t)]^{1/2} \xi(t), \quad (1)$$

where  $A(\omega_i, t)$  is a time dependent drift given by

$$A(\omega_i, t) = (\alpha + \lambda)W(t)\Pi_i - \lambda\omega_i - \beta\omega_0 \quad (2)$$

and the diffusion term by

$$D(\omega_i, t) = (\alpha + \lambda)W(t)\Pi_i + \lambda\omega_i + \beta\omega_0 - 2\lambda\omega_i\Pi_i. \quad (3)$$

Application of the Central Limit Theorem guaranties the convergence of the noise  $\xi(t)$  to a gaussian white noise in the limit  $W(t) \gg 1$ . The first term in the right hand side in Eq. (2) is a creation term accounting for new and old users that choose node  $i$  as a provider. The second term represent those users who decide to change their providers and, finally, the last term corresponds to the decrease of users due to introduction of newly created ASs. To proceed further, we need to specify the preference function  $\Pi_i(\{\omega_j(t)\})$ . We assume that, as a result of a competition process, bigger ASs get users more easily than small ones. The simplest function satisfying this condition corresponds to the linear preference, that is,

$$\Pi_i(\{\omega_j(t)\}) = \frac{\omega_i}{W(t)} \quad (4)$$

where  $W(t) = \omega_0 N_0 \exp(\alpha t)$ . In this case, the stochastic differential equation (1) reads

$$\frac{d\omega_i}{dt} = \alpha\omega_i - \beta\omega_0 + [(\alpha + 2\lambda)\omega_i + \beta\omega_0]^{1/2} \xi(t). \quad (5)$$

Notice that reallocation of users (*i.e.* the  $\lambda$ -term) only increases the diffusive part in Eq. (5) but has no net effect in the drift term, which is, eventually, the leading term. The complete solution of this problem requires to solve the Fokker-Plack equation corresponding to Eq. (5) with a reflecting boundary condition at  $\omega = \omega_0$  and initial conditions  $p(\omega_i, t_i | \omega_0, t_i) = \delta(\omega_i - \omega_0)$  ( $\delta(\cdot)$  stands for the Dirac delta function). Here  $p(\omega_i, t | \omega_0, t_i)$  is the probability that node  $i$  has wealth  $\omega_i$  at time  $t$  given that it had  $\omega_0$  at time  $t_i$ . The choice of a reflecting boundary condition at  $\omega = \omega_0$  is equivalent to assume that  $\beta$  is the overall growth rate of the number of nodes, that is, the composition of the birth and dead processes ruling the evolution of the number of nodes.

Finding the solution for this problem is not an easy task. Fortunately, we can take advantage of the fact that, when  $\alpha > \beta$ , the average number of users of each node increases exponentially and, since  $D(\omega_i, t) = \mathcal{O}(A(\omega_i, t))$ , fluctuations vanishes in the long time limit. Under this zero noise approximation, the number of hosts connected to an AS introduced at time  $t_i$  is

$$\omega_i(t | t_i) = \frac{\beta}{\alpha}\omega_0 + (1 - \frac{\beta}{\alpha})\omega_0 e^{\alpha(t-t_i)}. \quad (6)$$

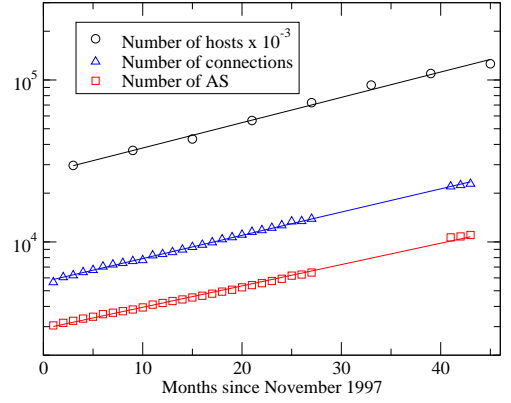


FIG. 1: Temporal evolution of the number of hosts, autonomous systems and connections among them from November 1997 to May 2002. Solid lines are the best fit estimates which gives the values for the rate growths of  $\alpha = 0.036$ ,  $\beta = 0.029$  and  $\delta = 0.033$  (units are month<sup>-1</sup>).

The probability density function of  $\omega$  can be calculated in the long time limit as

$$p(\omega, t) = \beta e^{-\beta t} \int_0^t e^{\beta t_i} \delta(\omega - \omega_i(t | t_i)) dt_i \quad (7)$$

which leads to

$$p(\omega, t) = \frac{\tau(1 - \tau)^\tau \omega_0^\tau}{(\omega - \tau\omega_0)^{1+\tau}} \Theta(\omega_c(t) - \omega), \quad (8)$$

where we have defined  $\tau \equiv \beta/\alpha$  and the cut-off is given by  $\omega_c(t) \sim (1 - \tau)\omega_0 e^{\alpha t} \sim W(t)$ . Thus, in the long time limit,  $p(\omega, t)$  approaches a stationary distribution with an increasing cut-off. In the case of the Internet,  $\alpha \gtrsim \beta$  which implies an exponent smaller but close to 2. A similar result was obtained in [9].

The key point in what follows is how to relate the number of users attached to an AS with its degree. Our basic assumption is that vertices are continuously adapting their bandwidth to the number of users they have. However, once an AS decides to increase its bandwidth it has to find a peer who, at the same time, wants to increase its bandwidth as well. The reason is that connection costs among ASs must be assumed by both peers. This fact differs from other growing models in which vertices do not ask target vertices if they really want to form those connections. Our model is, then, to be thought of as a coupling between a competition process for resources and adaptation of vertices to their current situation, with the constraint that connections are only formed between active nodes. Let  $b_i(t | t_i)$  be the total bandwidth of an AS at time  $t$  given that it was introduced at time  $t_i$ . This quantity can include single connections with other ASs, *i.e.* the topological degree  $k$ , but it also accounts for connections which have higher capacity. This is equivalent to say that the network is, in fact, weighted and  $b_i$  is the weighted degree. To simplify the model we consider

that bandwidth is discretized in such a way that single connections with high capacity are equivalent to multiple connections between the same ASs. Then, when a pair of ASs agrees to increase their mutual connectivity the connection is newly formed if they were not previously connected or, if they were, their mutual bandwidth increases by one unit. Now, we assume that, at time  $t$ , each AS adapts its total bandwidth proportionally to its number of users. We can write

$$b_i(t|t_i) = 1 + a(t) (\omega_i(t|t_i) - \omega_0). \quad (9)$$

Summing Eq. (9) for all nodes we get

$$a(t) = \frac{2B(t) - N(t)}{W(t) - \omega_0 N(t)} \sim \frac{2B(t)}{W(t)}, \quad (10)$$

where  $B(t)$  is the total bandwidth of the network which is, obviously, an upper bound to the total number of edges of the network. This suggests that  $B(t)$  will grow according to  $B(t) = B_0 e^{\delta' t}$ , where  $\delta' \gtrsim \delta$ . Using this assumption, we can express the individual bandwidth as  $b_i(t|t_i) \sim N(t)^{(\delta' - \alpha)/\beta} \omega_i(t|t_i)$ . From this equation, the scaling of the maximum bandwidth with the system size reads  $b_c(t) \sim N(t)^{\delta'/\beta}$ , that is, faster than  $N(t)$ . This implies that the network must necessarily contain multiple connections. Then, we propose that degree and bandwidth are related, in a statistical sense, through the following scaling relation

$$k(t|t_i) \sim [b(t|t_i)]^\mu, \quad (11)$$

where the scaling exponent,  $\mu < 1$ , is obtained by imposing that the maximum degree scales linearly with  $N(t)$  [19]. This sets the scaling exponent to  $\mu = \beta/\delta'$ . All four growth rates in the model are not independent but can be related by exploring the interplay between bandwidth, connectivity and traffic of the network. As the number of users grow, the global traffic of the Internet also grows, which means that ASs do not only adapt their bandwidth to their number of users but to the global traffic of the network. Therefore,  $a(t)$  must be an increasing function of  $t$ , which, in turn, implies that  $\delta' > \alpha$ . Using this condition and summing Eq. (11) for all vertices, the scaling of the total number of connections is  $E(t) \sim N(t)^{2-\alpha/\delta'}$ , which leads to  $\delta' = \alpha\beta/(2\beta - \delta)$ . Combining this relation with Eqs. (8), (9) and (11), the degree distribution reads

$$P(k) \sim \frac{\tau(1-\tau)^\tau [\omega_0 a(t)]^\tau}{\mu} \frac{1}{k^\gamma} \Theta(k_c(t) - k), \quad (12)$$

where the exponent  $\gamma$  takes the value

$$\gamma = 1 + \frac{\beta}{2\beta - \delta}. \quad (13)$$

Strikingly, the exponent  $\gamma$  has lost any dependence on  $\alpha$  becoming a function of the growth rate of both the number of ASs and the number of connections of the network.

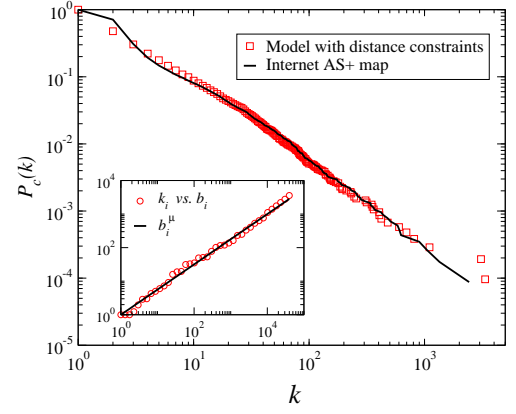


FIG. 2: Cumulative degree distribution ( $P_c(k) = \sum_k P(k')$ ) for the extended AS map compared to simulations of the model,  $r = 0.8$ . Inset: AS's degree as a function of AS's bandwidth. The solid line stands for the scaling relation Eq. (11) with  $\mu = \beta/\delta' = 0.75$ .

Using the empirical values for  $\beta$  and  $\delta$ , the predicted exponent is  $\gamma = 2.2 \pm 0.1$ , in excellent agreement with the values reported in the literature [11, 12, 13].

So far, we have been mainly interested in the degree distribution of the AS map but not in the specific way in which the network is formed. To fill this gap we have performed numerical simulations that generate network topologies in nice agreement with real measures of the Internet that go beyond the degree distribution. We consider a realistic geographical deployment of ASs and physical distance among them to take into account connection costs [17]. Our algorithm, following the lines of the model, works in four steps:

1. At iteration  $t$ ,  $\Delta W(t) = \omega_0 N_0 (e^{\alpha t} - e^{\alpha(t-1)})$  users join the network and choose provider among the existing nodes using the preference rule Eq. (4).
2.  $\Delta N(t) = N_0 (e^{\beta t} - e^{\beta(t-1)})$  new ASs are introduced with  $\omega_0$  users each, those being randomly withdrawn from already existing ASs. Newly created ASs are located in a two dimensional plane following a fractal set of dimension  $D_f = 1.5$  [17].
3. Each AS evaluate its increase of bandwidth,  $\Delta b_i(t|t_i)$ , according to Eq. (9).
4. A pair of nodes,  $(i, j)$ , is chosen with probability proportional to  $\Delta b_i(t|t_i)$  and  $\Delta b_j(t|t_j)$  respectively, and, whenever they both need to increase their bandwidth, they form a connection with probability  $D(d_{ij}, \omega_i, \omega_j)$ . This function takes into consideration that, due to connection costs, physical links over long distances are unlikely to be created by small peers. Once the first connection has been formed, they create a new connection with probability  $r$ , whenever they still need to increase their

bandwidth. This step is repeated until all nodes have the desired bandwidth.

It is important to stress the fact that nodes must be chosen with probability proportional to their increase in bandwidth at each step. The reason is that those nodes that need a high bandwidth increase will be more active when looking for partners to whom form connections. Another important point is the role of the parameter  $r$ . This parameter takes into account the balance between the costs of forming connections with new peers and the need for diversification in the number of partners. The effect of  $r$  in the network topology is to tune the average degree and the clustering coefficient by creating more multiple connections. The exponent of the degree distribution is unaffected except in the limiting case  $r \rightarrow 1$ . In this situation, big peers will create a huge amount of multiple connections among them, reducing, thus, the maximum degree of the network. Finally, we chose an exponential form for the distance probability function  $D(d_{ij}, \omega_i, \omega_j) = e^{-d_{ij}/d_c(\omega_i, \omega_j)}$ , where  $d_c(\omega_i, \omega_j) = \omega_i \omega_j / \kappa W(t)$  and  $\kappa$  is a cost function of number of users per unit distance, depending on the maximum distance of the fractal set. All simulations are performed using  $\omega_0 = 5000$ ,  $N_0 = 2$ ,  $B_0 = 1$ ,  $\alpha = 0,035$ ,  $\beta = 0,03$ , and  $\delta' = 0,04$ , and the final size of the networks is  $N \sim 11000$ . Simulations will be compared to the AS+ extended map recorded on May 2001, as reported in [18] that offers a better picture of the actual map.

Fig. 2 shows simulation results for the cumulative degree distribution, in nice agreement to that measured for the AS+ map. The inset exhibits simulation results of the AS's degree as a function of the AS's bandwidth, conforming to the scaling relation in Eq. (11). Clustering coefficient and average nearest neighbors degree are showed in Fig. 3. Dashed lines result from the model without distance constraints, whereas squares correspond to the model with distance constraints. Interestingly, the high level of clustering coming out from the model arises as a consequence of the pattern followed to attach nodes, so that only those AS willing for new connections will link. As can be observed in the figures, distance constraints introduce a disassortative component by inhibiting connections between small ASs so that the hierarchical structure of the real network is better reproduced.

We conclude by pointing out that this work is a first attempt towards a more realistic and complete modeling of the Internet, which, for instance, is of utmost importance in new communication protocols testing, which can be very sensitive to topological details. We would like to stress that the relevance of our model resides in the robustness of a simple statistical physics approach and, as a result, the unprecedented completeness of the topological description of the Internet and the novel insights into the dynamical processes leading to network formation.

We acknowledge R. Pastor-Satorras and A. Arenas for

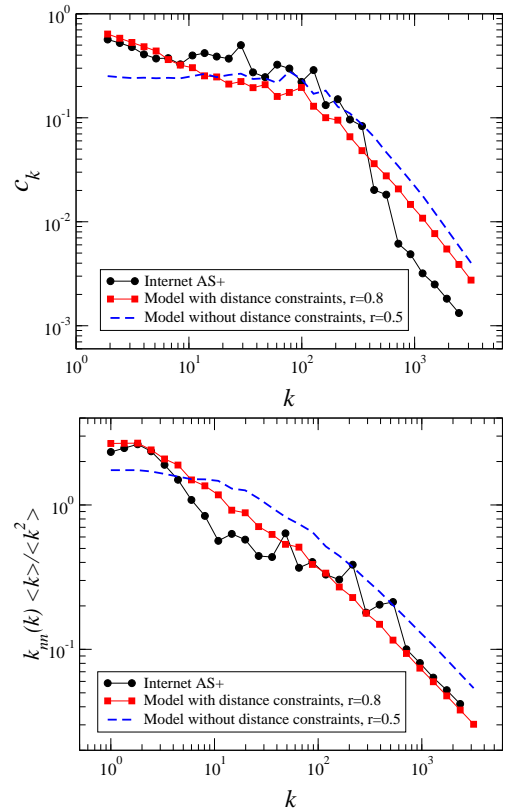


FIG. 3: Clustering coefficient,  $c_k$  (top) and normalised average nearest neighbors degree,  $\bar{k}_{nn}(k)\langle k \rangle / \langle k^2 \rangle$  (bottom), as functions of the node's degree for the extended autonomous system map (circles) and for the model with and without distance constraints (red squares and dashed line, respectively).

valuable comments and suggestions. This work has been partially supported by DGES of the Spanish government, Grant No. BFM2003-08258, and EC-FET Open project COSIN IST-2001-33555. M. B. acknowledges financial support from the MCyT (Spain).

- 
- [1] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet. A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004).
  - [2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
  - [3] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
  - [4] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
  - [5] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
  - [6] B. A. Huberman and L. A. Adamic, *Nature (London)* **401**, 131 (1999).
  - [7] K.-I. Goh, B. Kahng, and D. Kim, *Phys. Rev. Lett.* **88**, 108701 (2002).
  - [8] A. Capocci, G. Caldarelli, R. Marchetti, and L.

- Pietronero, Phys. Rev. E **64**, 035105 (2001).
- [9] M. Fayed, P. Krapivsky, J. W. Byers, M. Crovella, D. Finkel, and S. Redner, Comput. Commun. Rev. **33**, 41 (2003).
  - [10] S. Zhou, and R. J. Mondragon, cs.NI/0402011.
  - [11] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
  - [12] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
  - [13] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
  - [14] Data from the Hobbes' Internet Timeline <http://www.zakon.org/robert/internet/timeline/>.
  - [15] Available data correspond to daily measurements from November 1997 to January 2000 and three months of 2002. <http://moat.nlanr.net/Routing/rawdata/>
  - [16] S.N. Dorogovtsev, and J.F.F. Mendes, Phys. Rev. E **63**, 025101 (2001).
  - [17] S. H. Yook, H. Jeong, and A. -L. Barabási, Proc. Nat. Acad. Sci. USA **99**, 13382 (2002).
  - [18] C. Qian, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, Proceedings IEEE, IEEE Computer Society Press **2**, 608 (2002).
  - [19] Empirical measurements made in [7] showed such linear scaling in the AS with the largest degree.